SciDAC DataGrid Middleware
A High-Performance Data Grid Toolkit:
Enabling Technology for Wide Area Data-Intensive Applications
Quarterly Report April 2002 thru June 2002

Accomplishments this Quarter:

Continued development and testing of alpha prototype of the Giggle Replica Location Service
We continued extensive efforts at further development and testing of the Replica Location Service. Features added to the RLS this quarter include: support for user-defined attributes, including master copy attributes required by PPDG; alternatives for updating Replica Location Index nodes that include complete soft state updates and incremental updates; a simple partitioning scheme for Replica Location Index nodes based on pattern-matching of logical file names; the use of redundant Replica Location Index nodes for fault tolerance and load balancing. Many of our efforts this quarter were devoted to making the code more robust. We also made some modifications to the API. In addition to our functional and performance testing, some functional testing of the RLS was performed by the WP2 group of the European DataGrid project.

General release of alpha code for Giggle Replica Location Service
In June 2002, we released our latest alpha code for general release and testing by the grid community. The code and documentation for the RLS are available at: http://www.isi.edu/~annc/RLS.html

Designed a Metadata Catalog Service and began implementation
During the last quarter, we designed the schema and API for a simple Metadata Catalog Service. This service will maintain metadata that describes the contents of logical files in a data grid. The Metadata Catalog Service (MCS) allows users to query based on attributes of data rather than data names. In addition, the MCS provides management of logical collections and logical views of files. The MCS also includes support for use of containers, which are aggregations of small files that are stored, moved and replicated together. We have begun implementation of the MCS, studying alternatives such as the Spitfire database service and our own relational database service implementation adapted from the RLS implementation. We have released the Metadata Catalog Service schema document to the Earth Systems Grid group and other applications and are currently collecting feedback on the design.

Hardened existing grid software.

- Redesigned the Globus GASS Cache, yielding a speedup of two orders of magnitude.
- Subjected research systems such as Kangaroo and NeST to cluster-scale deployment in order to eliminate bugs and understand manageability issues.
- Continued development and periodic release of the public ClassAd resource management library (http://www.cs.wisc.edu/condor/classad) Released a soft-state cataloging system for resource advertising and discovery via ClassAds. (http://www.cs.wisc.edu/condor/catalog)

Continued development of the Kangaroo I/O system. The Kangaroo I/O system aims to harness all available bandwidth and buffering capacity of a data grid when moving data to and from executing jobs.

By making large amounts of buffers space transparently accessible, jobs may be insulated from the inevitable performance variations and network outages on the grid. This quarter, we developed:

- A variety of interfaces that each specialize a storage device to the task at hand: archival, caching, buffering, or process synchronization.
- A system for pipelining simultaneous processes using the phylum of storage interfaces described above.
- The triggering of I/O operations on CPU completion and vice-versa.

Continued development of an error handling framework for grid computing. Complex distributed systems fail frequently and in expected ways. A layered, multi-vendor, distributed I/O system will be extraordinarily sensitive to errors as failures cross vendor boundaries. This effort aims to improve the understanding and handling of errors through both formal techniques and practically deployable tools. This quarter, we developed:

- A formal statement of error structures, analysis, and advice for system builders.
- A procedural scripting language (FTSH - Fault Tolerant Shell) for robustly executing programs in the face of errors with strict user controls on resource consumption.
- Modifications to a declarative language (DAGMan) for managing failures in a similar fashion

Continued development of the NeST Storage Appliance. NeST is a flexible, software-only storage appliance which is designed for ease of configuration and deployment in a grid environment. This quarter, we developed:

- "Best-effort lots," relaxed allocations of space which allow resource owners priority without denying opportunistic use to resource borrowers.
- An NFS protocol handler to allow legacy systems to interact with grid-aware appliances.
- Different scheduling mechanisms, both to increase performance using cache-aware scheduling and to improve usability by exposing configurable proportional bandwidth sharing using stride scheduling.

Work on Callback Spaces was completed
This was un-scheduled work. This solved a problem for the Grads Project in that it allowed threaded application code to call non-threaded Globus code without conflict. It also had the side benefit (to GridFTP) of allowing multiple threads to handle callbacks and thus potentially having significant performance benefits on IO in threaded builds of GridFTP.

Extensive code review for timing related bugs
Several bugs were reported that were as a result of subtle race conditions. The callback spaces work referenced above also surfaced several related bugs. We therefore decided to suspend further development and conduct an extensive code review. This resulted in several bugs being fixed and an overall more stable GridFTP server.

Reliable File Transfer Service
The RFT is scheduled to be the first OGSA compliant service. Modifications continue to be made to keep the implementation compliant with the OGSA specification as it evolves. Additional features have been added including security integration and ability to control more performance attributes.

Plans for next quarter

Demo RFT at GGF/HPDC
RFT will be one of the prime features in a GT3/OGSA demo to be held at GGF/HPDC in Edinburgh Scotland in July.

Continued testing and development of alpha prototype of Replica Location Service
We will continue to work on functional, integration, stress and performance testing of the RLS prototype. We hope to extend the set of alpha testers to include users from PPDG, ESG and GriPhyN communities. Functionality still to be added to the RLS includes integration with the Community Authorization Service and development of a SOAP/OGSA interface.

Release of the RLS into the Globus Toolkit
We plan to migrate a beta version of the prototype RLS code into the Globus Toolkit in the coming quarter.

Produce alpha prototype of Metadata Service
Based on feedback from those who have reviewed the schema and API documents for the Metadata Catalog Service, we will implement an alpha prototype for the service and begin testing it, most likely using metadata from the Earth Systems Grid project.

Execute a feasibility study for integration of GridFTP into HPSS
This was planned for last quarter, but was delayed due to the callback spaces work. We will be working with Los Alamos National Laboratory and the HPSS development team to assess the feasibility of replacing the HPSS data mover protocol with GridFTP. The proposal is that this will make it into production HPSS, making all HPSS systems natively accessible by GridFTP clients.

Release an alpha of the GridFTP V1.1 that includes usability enhancements
This was planned for last quarter, but was delayed due to the callback spaces work. We are currently working on a number of enhancements to increase usability of the server and make support easier. These include runtime version dumps to verify versions of all components including ones that are dynamically linked, options to control the frequency of restart and performance markers, and support for connection decisions to be based on filenames to support proxy applications.

Design for Server Development
We are going to be doing a light weight, from scratch server development. This primarily to support the ability to start up a striped server from within a parallel application running on a scheduled resource where you can not pre-install servers. This will also allow us to bring the striped and non-striped code bases back together.

Design and alpha prototype development of Replica Management Service
We will reach consensus on our design proposal for the Replica Management Service and implement and test an alpha prototype of the service.

Design and alpha prototype development of Metadata Service
We will implement a prototype of a simple, centralized metadata service. This may be based on the Spitfire database service from the European DataGrid project or may be based on our own server in front of a mySQL database.

Papers Published or in Progress

Giggle: A Framework for Constructing Scalable Replica Location Services. Ann Chervenak, Ewa Deelman, Ian Foster, Leanne Guy, Wolfgang Hoschek, Adriana Iamnitchi, Carl Kesselman, Peter Kunszt, Matei Ripeanu, Bob Schwartzkopf, Heinz Stockinger, Kurt Stockinger, Brian Tierney. Accepted for SC2002 conference.

Douglas Thain and Miron Livny, "Error Scope on a Computational Grid: Theory and Practice," to appear in Proceedings of the Eleventh IEEE Symposium on High Performance Distributed Computing, HPDC-11, Edinburgh, Scotland, July 2002. (http://www.cs.wisc.edu/condor/doc/error-scope.pdf)

John Bent, Venkateshwaran Venkataramani, Nick LeRoy, Alain Roy, Joseph Stanley, Andrea C. Arpaci Dusseau, Remzi H. Arpaci-Dusseau and Miron Livny, "Flexibility, Manageability, and Performance in a Grid Storage Appliance," to appear in Proceedings of the Eleventh IEEE Symposium on High Performance Distributed Computing, HPDC-11, Edinburgh, Scotland, July 2002. (http://www.cs.wisc.edu/condor/nest/papers/nest-hpdc-02.pdf)

Nathan Burnett, John Bent, Andrea Arpaci-Dusseau, Remzi Arpaci-Dusseau, "Exploiting Gray-Box Knowledge of Buffer-Cache Management," The 2002 USENIX Annual Technical Conference, Monterey, California, June 2002. (http://www.cs.wisc.edu/wind/Publications/dust-usenix02.pdf)

Presentations Given

1. **Douglas Thain, "Error Scope on a Computational Grid"**, Paradyn/Condor Week 2002.
2. **John Bent, "NeST: A status report",** Paradyn/Condor Week 2002.